DOCUMENT RESUME

ED 064 359 TM 001 527

AUTHOR Calkins, Dick S.; Jennings, Earl

TITLE An Empirical Investigation of Some Effects of the Violation of the Assumption That the Covariable in Analysis of Covariance Is a Mathematical Variable.

NOTE 29p.; Paper is partially based on the doctorial dissertation of D. S. Calkins

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS *Analysis of Covariance; *Factor Analysis; *Linear Programing; Mathematical Applications; *Mathematical Models; *Probability; Statistical Analysis

ABSTRACT

The mathematical derivation of the statistics used for inference in some linear models assumes that values of the independent variables are pre-selected such that these variables can be treated as fixed rather than random variables. This assumption is often disregarded when these models are utilized in research. This study is an investigation of the consequences of the violation of this assumption. The results of this study indicate that when the sample size is not too small the consequences of the violation of this assumption are of little practical significance. (Author)

AN EMPIRICAL INVESTIGATION OF SOME EFFECTS OF THE VIOLATION OF THE ASSUMPTION THAT THE COVARIABLE IN ANALYSIS OF COVARIANCE IS A MATHEMATICAL VARIABLE

Dick S. Calkins
The University of Texas at El Paso
and Earl Jennings
The University of Texas at Austin

Abstract. The mathematical derivation of the statistics used for inference in some linear models assumes that values of the independent variables are pre-selected such that these variables can be treated as fixed rather than random variables. This assumption is often disregarded when these models are utilized in research. This study is an investigation of the consequences of the violation of this assumption. The results of this study indicate that when the sample size is not too small the consequences of the violation of this assumption are of little practical significance.

AN EMPIRICAL INVESTIGATION OF SOME EFFECTS OF THE VIOLATION OF THE ASSUMPTION THAT THE COVARIABLE IN ANALYSIS OF COVARIANCE IS A MATHEMATICAL VARIABLE[1]

Dick S. Calkins
The University of Texas at El Paso
and Earl Jennings
The University of Texas at Austin[2]

In many cases the analysis of data in behavioral research can be accomplished through the formulation of a linear model which appears to represent the essential aspects of a suspected relationship between the independent and dependent variables being investigated. In such a model the data appear as variables and the statistics appear as constants which are calculated from the data. When certain mathematical procedures are used for calculating the values of the statistics and certain assumptions have been met concerning the selection and distribution of the values of the variables in the population from which the data were drawn, it can be shown mathematically that the statistics are "good" estimates of the parameters and that accurate probability statements involving possible differences in the parameters can be made. However, if the assumptions concerning the selection and distribution of the values of the variables in the population are not met, it may be very difficult to show mathematically how the estimates of the parameters and the probability statements involving differences in the parameters will be affected. In some instances, the

[1]This paper is partially based on the Doctoral Dissertation of the first author. The second author served as chairman of that dissertation committee.

[2]Dick S. Calkins is an Assistant Professor in the Department of Educational Psychology and Guidance and a Research Consultant for the Computation Center at The University of Texas at El Paso. Earl Jennings is an Associate Professor in the Department of Educational Psychology and Assistant Director of the Measurement and Evaluation Center at The University of Texas at Austin.

effects of the violation of these assumptions may be of small enough magnitude that they are of no practical significance. The magnitude of the effects of the violation of such assumptions can be investigated empirically by repeatedly sampling from populations of values with known characteristics when the assumptions to be investigated are not met. This study was performed to investigate the effects of the violation of one such assumption.

For the mathematical model which is the basis for regression analysis, it is necessary that the continuous independent variables are mathematical variables (in contrast to random variables) before it can be shown that the statistics are "good" estimates of the parameters or that the shape of the sampling distribution of the statistics follows the normal distribution (Graybill, 1961, pp. 195-200, 383-396). If the sampling distributions of the statistics are non-normal, the probability statements involving differences in the parameters may be inaccurate. The intent of this study was to investigate the effects produced for a particular family of linear models which contain both continuous and binary coded independent variables when the assumption that the continuous variables are mathematical is not maintained.

The Models

The linear models investigated can be utilized to alleviate a frequently occurring problem in behavioral research. This problem arises when an investigator desires to examine differences in existing groups where the differences could be attributable to some quantifiable concomitant influence. In such a situation, the investigator would probably want to investigate possible differences in the performance of the groups as measured by some dependent variable without regard to differences due to the concomitant variable. A logical approach to this difficulty would be to consider the joint frequency distribution for the dependent variable and the concomitant variable for each group. Comparison of the joint frequency distributions for the groups in effect makes possible the comparison of values of the dependent variable for individuals in the various groups who have the same value for the concomitant variable.

Bottenberg and Ward (1963, pp. 76-86) present a family of linear models which can be used to make these previously mentioned comparisons in a more quantitative manner. The most general model of this family is written

$$Y_{ij} = a_j + b_j X_{ij} + e_{ij} \qquad \text{Model 1}$$

where Y_{ij} is the value of the dependent variable, X_{ij} is the value of the concomitant variable, and e_{ij} is the error associated with use of Model 1 with the ith member of the jth group; $j = 1, 2, \ldots, m$ where m is the number of groups and $i = 1, 2, \ldots, n_j$ where n_j is the number of individuals in the jth group. (In this model it is assumed that within any populations from which the groups were selected, the expected change in the value of the dependent variable per unit change in the value of the concomitant variable is constant over the range of the values of the concomitant variable.) The evaluation of the constants in this model from the data produces a unique value of a and b for each group. The value of a and b for each group results from fitting a regression line to the joint frequency distribution for the dependent variable and the concomitant variable for each group. The values of a and b then represent the intercept and slope of the regression line for each group. The determination of whether the various groups differ on the dependent variable without regard to differences due to the concomitant variable can be made in terms of the intercepts and slopes of the group regression lines.

Probability statements involving possible differences in the intercept and the slope parameters for the populations from which the groups were selected can be made by calculation of a critical statistic which is a function of the error sum of squares in Model 1 and in models derived in particular ways from Model 1. Probability statements involving differences in the slope parameters for the populations from which the groups were selected can be made on the basis of the value of a critical ratio which is a function of the error sum of squares (s) from Model 1 which can be written

$$s = \sum_{j=1}^{m} \sum_{i=1}^{n_j} e_{ij}^2$$

and the error sum of squares from a model derived from Model 1 which restricts all of the values of group slope to the same value. This restricted model is written

$$Y_{ij} = a_j + bX_{ij} + f_{ij} \qquad \text{Model 2}$$

where Y_{ij} is the value of the dependent variable, X_{ij} is the value of the concomitant variable, and e_{ij} is the error associated with the use of Model 2 with the *i*th member of the *j*th group; $j = 1, 2, \ldots, m$ where *m* is the number of groups and $i = 1, 2, \ldots, n_j$ where n_j is the number of individuals in the *j*th group. The common value of group slope is *b* and the *a*'s are the group intercepts. The error sum of squares (t) from Model 2 can be written

$$t = \sum_{j=1}^{m} \sum_{i=1}^{n_j} f_{ij}^2.$$

Probability statements involving possible differences in the intercept parameters for the populations from which the groups were selected, assuming that the slope parameters for these populations are equal, can be made on the basis of the value of a critical statistic which is a function of the error sum of squares (t) from Model 2 and the error sum of squares (r) from a model derived from Model 2 which restricts all the values of group intercept to the same value. This restricted model is written

$$Y_{ij} = a + bX_{ij} + g_{ij} \qquad \text{Model 3}$$

where Y_{ij} is the value of the dependent variable, X_{ij} is the value of the concomitant variable, and e_{ij} is the error associated with the use of Model 3 with the *i*th member of the *j*th group; $j = 1, 2, \ldots, m$ where *m* is the number of groups and $i = 1, 2, \ldots, n_j$ where n_j is the number of individuals in the *j*th group. In Model 3, *a* is the value of the common intercept for all the groups and *b* is the value of the common slope for all the groups. The error sum of squares (r) from Model 3 can be written

$$r = \sum_{j=1}^{m} \sum_{i=1}^{n_j} g_{ij}^2 .$$

The extent of this research was to investigate the properties of Models 1, 2 and 3 and the probability statements based on these models when the concomitant variable was not a mathematical variable. The comparisons made possible by the use of these models are essentially those made in analysis of covariance.

Mathematical and Random Variables

In the mathematical treatment of a linear model which is necessary in order to derive computing expression for the constants and the critical statistics, an important consideration is the type of variables which are used in the model. The two types of variables generally recognized by mathematical statisticians are random variables and mathematical variables. The following definition of a random variable has been adapted from Alexander (1961).

> If for a particular random experiment, $\{A_1, . . ., A_k\}$ is the set of outcomes (sample points) defining the sample space of the random experiment, and if $\{X_1, . . ., X_k\}$ is a set of numbers such that X_i is associated with the corresponding outcome A_i for $i = 1, . . ., k$ then the set of values $\{X_1, . . ., X_k\}$ is called a random variable for the particular random experiment.

For a variable to be considered a mathematical or fixed variable, the values assumed by the variable must be known constants; that is, the values of a mathematical variable must be pre-selected from the range of possible values assumed by a random variable. For example, consider the case where an investigator is interested in evaluating the relative effects of an experimental curriculum and a control curriculum after the removal of the unwanted influence of difference in initial performance. If the investigator desired to treat his assessment of initial performance as a fixed variable in Models 1, 2 and 3, it would be necessary for him to select certain values of initial performance before he tested the pupils and then to use in the analysis only the pupils who had those specific values of initial performance. If he wished to treat initial performance as a random variable in the models, he could simply assess the initial performance of all the available pupils and use their scores regardless of particular values.

Another important consideration involving the variables which appear in a linear model is concerned with the amount of error inherent in the process of observation of the values of the variable. Whether a variable is treated as fixed or random, the process by which the values of the variables are observed usually introduces some error of measurement. The relative magnitude of the error introduced by the measurement process is generally used a priori to determine whether the observed values of a variable will be treated in the model as measured with error or as error free. For example, it would probably not be reasonable to conclude that observed initial performance is an errorless assessment of a given subject's potential.

The mathematical derivation of the sampling distributions of the estimates of the unknown parameters and the properties of the critical ratios for linear models involve both the type of the variables and whether or not the variables are considered to be measured with or without error. In the mathematical treatment of models of the same type that are considered here, it is assumed that the X's are both fixed and measured without error (Graybill, 1961, pp. 103-104 and 383). Berkson (1950) has shown mathematically that if the X values are fixed variables but measured with error, the probability statements based on the critical ratio and the sampling distributions of the statistics are not effected. In consideration of further comments by Graybill associated with the assumptions underlying the various models and the associated mathematical development of the models which do consider various cases where the X's are treated as random variables both with and without error, it becomes apparent that a general solution to this problem is both difficult and unavailable.

There are instances in the natural and behavioral sciences when it is no problem to design experiments such that a concomitant variable is a fixed variable measured with very little error. For instance, if temperature were considered to be a variable which was critically affecting the comparison of yield in two or more manufacturing processes, all the processes could be utilized a given number of times at pre-selected values of temperature and then differences in the yields of the processes could be evaluated using Models 1, 2 and 3 to make possible comparison of the yield of the process with the effect of temperature removed. In the social sciences, if practice in an experiment concerning the effects of reinforcement on performance were thought to influence the comparison of performance for the various reinforcement conditions, there would be little problem involved in selecting certain amounts of practice and then assessing performance for a certain number of individuals for each reinforcement condition at the selected amounts of practice. With amount of practice a fixed variable measured with little error, use of Models 1, 2 and 3 to compare the performance of the various reinforcement conditions with the effect of practice removed is in correspondence with the assumptions for these models.

Unfortunately, the use of Models 1, 2 and 3 in the educational example when the concomitant variable is fixed is not a very satisfactory procedure because of the related problems of obtaining sufficient subjects who have the necessary scores on the concomitant variable within the constraints of the experimental situation. This problem is compounded by the difficulty involved in obtaining relatively errorless values of the concomitant variable. What usually occurs in actual practice is that the investigator ignores the requirement that the concomitant variable be fixed and error free and proceeds with the analysis as if this variable were fixed. The purpose of this study was to investigate the effects of the failure to meet these assumptions of the model when the X values are values of a random variable and are measured without error. This was accomplished by determining whether differences in the number of incorrect decisions based on the critical statistic for differences in the b's in Model 1 and the critical statistic for differences in the a's in Model 2 occur when X is a random variable rather than a mathematical variable. Also, comparisons were made between the distributions of the b's in Model 1 and of the a's in Model 2 when the X values were values of a random variable and values of a mathematical variable. The case when the X values represent values of a random variable measured with error was not treated in this study.

Methods

In order to conduct this empirical investigation, computer programs were written in FORTRAN to be run on the CDC 6600 Computer System at The University of Texas at Austin. These computer programs, which are shown in Calkins (1971), allowed values of Y to be randomly selected from various bivariate populations having predetermined parameters for values of X either fixed or randomly obtained. The specified characteristics of the bivariate populations were the type of bivariate distribution and the means and variances of the X and Y marginal distributions. (In actuality, it was difficult to maintain constant variance when the X values were fixed.) Factors of the investigation which were varied are number of cases (X,Y pairs) sampled from each group, shape of the X marginal distribution from which values of X were selected, and the variance of the Y values in each X array. The principal statistics which were observed are the distributions and expected values of the _a_'s and _b_'s, and the critical statistics based on possible differences in the _a_'s and _b_'s.

In general, the X and Y marginal distributions of the bivariate frequency distributions had means of 50.0 and standard deviations of 10.0, and correlations of 0.15, 0.30, 0.45, 0.60, 0.75 and 0.90 were used to determine the variance of the Y values for each X array. Samples of size five, 13 and 39 were used in experiments of 1,000 samples. The shapes of the X marginal distributions which were used are normal distributions and rectangular distribtuions. The corresponding types of bivariate distributions which were used are bivariate normal and the values of Y normally distributed for each value of X used but with all the Y arrays having equal variance.

Measurement of the Effects

Critical features of the various sampling distributions of the statistics were used to compare the effects of fixed and random selection of X values. The distributions of the _a_'s and the _b_'s were compared with their counterparts through the use of functions of the first four cumulative moments of their respective distributions. These statistics -- mean, variance, skewness and kurtosis -- were calculated using the computing expressions from Fisher (1958). It was expected that these statistics would closely approximate the population values.

The _F_ statistic is the critical statistic which was investigated. The _F_ statistic upon which decisions concerning possible differences in the _b_'s or slopes in Model 1 was denoted F_b, and the _F_ statistic upon which decisions concerning possible differences in the _a_'s or intercepts in Model 2 was denoted F_a. A concise presentation of a procedure for the caluculation of values of these statistics was adapted from Bottenberg and Ward (1963, pp. 76-86), although for the actual computations of these values, computing expressions from Winer (1962, pp. 578-588) were used.

$$F_b = \frac{(q_2 - q_1) / df_1}{q_1 / df_2} \quad \text{where } q_1 = \sum_j^m \sum_i^{n_j} e_{ij}^2$$

$$q_2 = \sum_j^m \sum_i^{n_j} f_{ij}^2$$

$$df_1 = m - 1$$

$$df_2 = \sum_{j=1}^{m} (n_j - 2)$$

e_{ij} is the error from Model 1 associated with individual *i* from group *j*

and f_{ij} is the error from Model 2 associated with individual *i* from group *j*.

$$F_a = \frac{(q_3 - q_2) / df_3}{q_2 / df_4} \quad \text{where } q_3 = \sum_j^m \sum_i^{n_j} g_{ij}^2$$

$$df_3 = m - 1$$

$$df_4 = \sum_{j=1}^{m} n_j - m - 1$$

and g_{ij} is the error from Model 3 associated with individual *i* from group *j*.

Since in this study all the samples for the groups were drawn from the same population, the expected values of F_a should be near $df_2/(df_2 - 2)$ and F_b should be near $df_4/(df_4 - 2)$. Also, not more than five percent of the values calculated for F_a and F_b should be equal to or greater than the specific values of the central *F* distribution for the proper degrees of freedom at the .05 confidence level. Departure from what is expected for either of these criteria would indicate that the values of the critical statistic are not *F* distributed, although the latter check is the more important, since departure from what is expected would upset the decision rule.

Sampling Procedure

In Monte Carlo studies such as this one, where sampling must be done from a joint frequency distribution representing a population of individuals, problems concerning efficient computer usage arise because large blocks of computer storage would be required to contain the frequency distribution. For this reason and other aspects of efficiency, actual frequency distributions were not used in this study. Instead of actually sampling from existing frequency distributions, random deviates were generated using computer programs based on pseudo random numbers such that these random deviates simulated sampling with replacement from distributions with desired characteristics.

The source of pseudo random numbers for this study was RANF[3], a FORTRAN function, which is available through the CDC 6600 computer system and documented in the computation center User's Manual of The University of Texas at Austin. The algorithm by which these pseudo random numbers were generated appears sufficient, for the purposes of this study, to consider the pseudo random numbers to be random. This function was utilized such that the same sequence of random numbers was used in each experiment.

The random numbers generated by RANF were used in two other functions, RNORMD and RANREC, to generate numbers which were random deviates of a normally distributed variable with a specified mean and variance in the case of RNORMD and random deviates from a rectangularly distributed variable with a specified mean and variance in the case of RANREC.

---

[3]Actually these so-called pseudo random numbers from RANF can be viewed as random samples from a continuous rectangular distribution which is defined only over the range zero to one. For purposes of this study deviates are defined to be random samples from distributions with specified characteristics which differ from the characteristics of the distribution inherent in RANF. Thus the numbers obtained from RANF are called random numbers and all numbers used in this study which are functions of the numbers obtained from RANF are called random deviates.

The random deviates from RNORMD and RANREC were used to produce numbers which were themselves random deviates from bivariate frequency distributions. (In the following discussion of the procedure used in the generation process, it may be helpful for the reader to refer to Figure 1.) Random deviates for a bivariate normal frequency distribution were generated by first obtaining a random normal value of X from a univariate distribution with a mean of 50.0 and a standard deviation of 10.0 by using RNORMD. The corresponding Y value is a random deviate from a normal distribution with a mean equal to the predicted value of Y for the particular value of X and a standard deviation which is the stadard error of the estimate ($S_{yx} = S_y\sqrt{1 - r_{yx}^2}$) for predicting Y from a knowledge of X. This random value of Y was thus generated by again using RNORMD to obtain a random normal deviate from a distribution with a mean equal to A + BX and a standard deviation equal to S_{yx} where X is the previously generated deviate and A, B and S_{yx} are values calculated from the parameters specified for the bivariate frequency distribution. This pair of X and Y values then represents a random deviate from a bivariate normal frequency distribution with specified X and Y univariate means and standard deviations and bivariate correlation.

For the case where it is necessary to obtain deviates from a bivariate normal frequency distribution for normally distributed but fixed values of X, the procedure for the generation of the Y values was the same as for generation of the random values of X but the procedure for obtaining the X values was different. Thirteen fixed values of X were chosen. These values were the mean of the X marginal distribution and six equally spaced values above and below this mean. The spacing of these values was determined in terms of the value of the standard deviation of the X marginal distribution such that these 12 values were equal to the mean plus or minus .5, 1.0, 1.5, 2.0, 2.5, and 3.0 times the standard deviation. The frequency of occurrence of each of these fixed values was used to determine the shape of the X marginal distribution. In order that the X marginal distribution be normally distributed, values of the probability function of the normal curve were obtained from a z score table using the X values in z score form as arguments. Since the height of the probability function

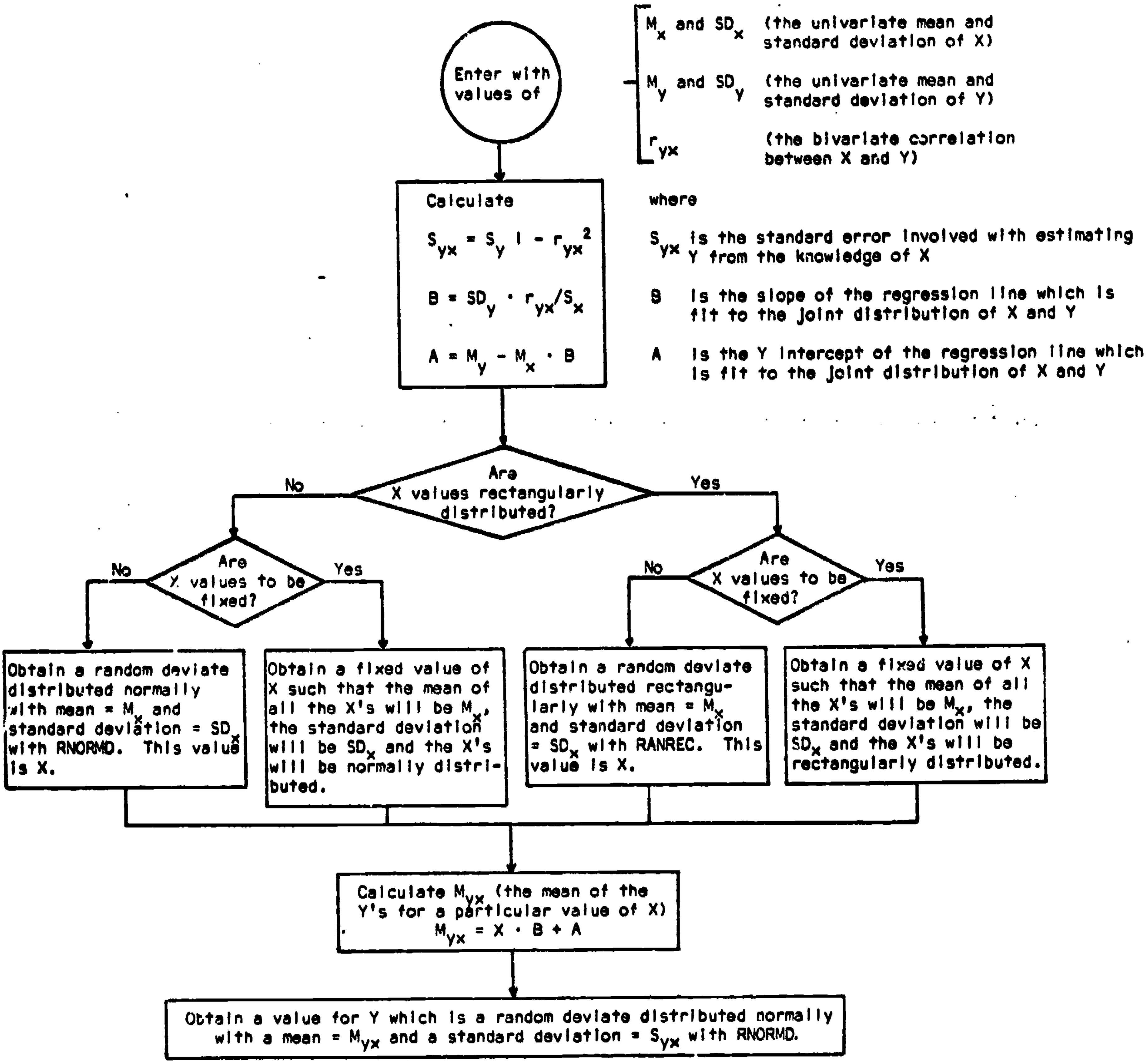

Figure 1. Flowchart showing the methods for obtaining the X and Y values for each of the various configurations of the bivariate frequency distributions.

of the normal curve for a particular z score can be interpreted as a proportion of the total number of cases occurring for that z score, the number of cases needed for any value of X for this particular selection of 13 z scores can be obtained by multiplying one-half the desired sample size by the height of the probability function of the normal curve for that X in z score units. The Y value corresponding to this value of X was then generated as in the random bivariate normal case. This pair of X and Y values then represents a deviate from a bivariate normal frequency distribution based on a fixed value of X.

For the case where it is necessary to obtain deviates from a bivariate frequency distribution for rectangularly distributed values of X but for the Y values normally distributed for each value of X, again the procedure for the generation of the Y values is the same as in the two previous cases. The random values of X were obtained by using RANREC to generate random deviates from a rectangular univariate distribution with a mean of 50.0 and a standard deviation of 10.0. The rectangularly distributed values of X were obtained in a manner analogous to the procedure previously described for obtaining normally distributed fixed values, except that in the rectangular case a rectangular probability function was utilized rather than the probability function for the normal curve.

However, it should be noted that when this procedure for both normal and rectangular distributions is used to establish the frequency of occurrence of the fixed X values, it is difficult to maintain both the system of intervals between the 13 fixed values of X and a given standard deviation of the X values. For this reason, in the fixed case the intervals between the fixed X values were maintained and the standard deviations of the X marginal distribution for different sample sizes were allowed to vary. For fixed but normally distributed values of X, the standard deviation of the X values was 7.07 for sample size five, 8.55 for sample size thirteen and 9.47 for sample size thirty-nine. For fixed but rectangularly distributed values of X, the standard deviation of the X values was 7.07 for sample size five, 18.71 for sample size thirteen and 18.71 for sample size thirty-nine.

The remainder of the procedure for all cases consisted of generating the necessary number of X,Y pairs with the appropriate characteristics for the desired number of cases per group, accumulating the various sums, sums of squares and sums of products and then utilizing these figures in the computing formulas to produce sample values of the slope, intercept, standard error of slope and intercept, and critical statistics for the slope and intercept. This entire procedure was then repeated one thousand times in order to obtain sampling distributions for the slope, intercept, and the critical statistics of the slope and intercept.

Results

The results are presented in Tables 2 through 7. Tables 2 and 3 represent the results of the experiments involving the investigation of the effects of violation of the assumption that the concomitant variable is a fixed or mathematical variable. Table 1 presents the legend necessary to interpret Tables 2 and 3. Tables 4 through 7 were prepared from the information contained in Tables 2 and 3 to aid in the interpretation of Tables 2 and 3.

Table 1 is an explication of the two and three letter codes which identify the various statistics reported for each experiment shown in Tables 2 and 3. It should be noted that the reported statistics for each experiment contain both expected and observed values pertaining to intercept and slope. The first set of five statistics refers to various expected and observed values of the distribution of slopes and the second set of five statistics refers to the same values of the distribution of intercepts. The next three statistics refer to various observed and expected values of the critical ratio related to differences in slope and the next three statistics refer to the same values except that they relate to differences in intercept.

Tables 2 and 3 show some of the expected and obtained values of the distributions of slope, intercept and critical statistics of the slope and intercept for the two types of bivariate distributions. The X's were selected with both fixed and random values for six values of standard errors of estimate based on the values of correlation shown for three sample sizes and two groups. Tables 4 through 7 contain summary information from Tables 2 and 3 concerning the discrepancy between the observed

Table 1

Legend of Alphabetic Codes Needed to Interpret Tables 3 through 7

ES - the expected mean of the theoretical sampling distribution of slope values which is calculated from the specified parameters by slope = $\frac{r \cdot S_y}{S_x}$

where r is the specified correlation
S_y is the standard deviation of the Y marginal distribution
S_x is the standard deviation of the X marginal distribution.

OS - the mean of the distribution of observed values of slope

SDS - the standard deviation of the distribution of the observed values of slope

SS - the skewness of the distribution of the observed values of slope

KS - the kurtosis of the distribution of the observed values of slope

EI - the expected mean of the theoretical sampling distribution of intercept values which is calculated from the specified parameters by intercept = M_x - slope$\cdot$ M_y

where M_x is the mean of the X marginal distribution
where M_y is the mean of the Y marginal distribution.

OI - the mean of the distribution of the observed values of intercept

SDI - the standard deviation of the distribution of the observed values of intercept

SI - the skewness of the distribution of the observed values of intercept

KI - the kurtosis of the distribution of the observed values of intercept

ODS - the number of observed F values based on differences in slope which are greater than the specified value of the central F distribution for the proper degrees of freedom at the .05/.01 confidence level

EFS - the expected mean of the central F distribution for the proper degrees of freedom for differences in slope

OFS - the mean of the observed distribution of F values based on difference in slope

ODI - the number of observed F values based on differences in intercept which are greater than the specified value of the central F distribution for the proper degrees of freedom at the .05/.01 confidence level.

EFI - the expected mean of the central F distribution for the proper degrees of freedom for difference in intercept

OFI - the mean of the observed distribution of F values based on differences in intercept

TABLE 2

Some Characteristics of the Sampling Distributions of Slope, Intercept and Critical Statistics of Slope and Intercept for Bivariate Normal Frequency Distributions for Random and Fixed Values of X for Six Values of Standard Error of Estimates and Three Sample Sizes for Two Groups

Number of Cases Per Group		Values of X obtained by Random Sampling From a Normal Distribution						Values of X fixed but Normally Distributed						
		Values of Standard Error of Estimate Based on the Correlations Shown												
		.1500	.3000	.4500	.6000	.7500	.9000	.1500	.3000	.4500	.6000	.7500	.9000	
	ES	.1500	.3000	.4500	.6000	.7500	.9000	.1500	.3000	.4500	.6000	.7500	.9000	ES
	OS	.1878	.3366	.4840	.6315	.7749	.9164	.1445	.2962	.4450	.5958	.7467	.8973	OS
	SDS	.6960	.6724	.6293	.5652	.5010	.3314	.6255	.6026	.5646	.5046	.4177	.2761	SDS
	SS	.5830	.6107	.6044	.6284	2.7825	2.8206	.0062	.0028	-.0018	.0020	.0010	.0125	SS
	KS	5.8524	6.0634	6.0055	6.2645	42.8697	43.8355	-.1652	-.1710	-.1651	-.1640	-.1658	-.1793	KS
	EI	42.5000	35.0000	27.5000	20.0009	12.5000	5.0000	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	EI
	OI	42.0914	34.5999	27.1589	19.6491	12.2444	4.8291	42.6794	35.0857	27.6191	20.1024	12.5754	5.0713	OI
	SDI	20.4849	20.1292	18.8517	16.8749	13.9488	9.2304	21.4648	21.1381	20.0281	17.9583	14.8982	9.9057	SDI
5	SI	-.1886	-.0892	-.0778	-.0586	-.0561	-.0644	-.0215	.0431	.0851	.0745	.0692	.0687	SI
	KI	.5452	.6394	.6572	.6596	.6804	.7091	-.1784	-.0673	.0489	.0393	-.0362	-.0436	KI
	ODS	55/7	52/8	51/8	53/8	51/9	53/8	48/11	51/11	51/11	49/11	54/12	51/12	ODS
	EFS	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	EFS
	OFS	1.5337	1.5359	1.5340	1.5325	1.5217	1.5261	1.7628	1.7743	1.7416	1.7931	1.7342	1.5884	OFS
	ODI	56/14	52/13	51/13	56/13	56/13	51/13	58/10	60/11	59/11	58/10	61/10	59/10	ODI
	EFI	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	EFI
	OFI	1.5080	1.5171	1.4918	1.4873	1.5188	1.5260	1.4812	1.4966	1.4930	1.4806	1.4845	1.4888	OFI
	ES	.1500	.3000	.4500	.6000	.7500	.9000	.1500	.3000	.4500	.6000	.7500	.9000	ES
	OS	.1539	.3035	.4538	.6036	.7520	.9017	.1356	.2862	.4366	.5882	.7405	.8937	OS
	SDS	.3207	.3091	.2892	.2593	.2146	.1416	.3206	.3095	.2897	.2599	.2147	.1419	SDS
	SS	-.0550	-.0402	-.0294	-.0408	-.0458	-.0411	.0296	.0310	.0305	.0330	.0362	.0389	SS
	KS	1.4193	1.2868	1.3917	1.3260	1.3149	1.4239	.0029	.0238	.0287	.0297	.0136	.0408	KS
	EI	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	EI
	OI	42.3395	34.8683	27.3679	19.8779	12.4408	4.9496	43.2188	35.6872	28.1610	20.5819	12.9608	5.3049	OI
	SDI	11.1438	10.7570	10.0527	9.0382	7.4901	4.9363	11.3989	11.0168	10.3471	9.3087	7.7297	5.1475	SDI
13	SI	.0365	.0375	.0356	.0396	.0555	.0517	-.0200	-.0213	-.0240	-.0327	-.0599	-.0236	SI
	KI	.2606	.2579	.2816	.2407	.2214	.1909	.2129	.1869	.1919	.1954	.0366	.2279	KI
	ODS	53/11	50/10	55/10	57/9	60/10	54/11	53/6	52/5	52/6	56/7	52/6	53/6	ODS
	EFS	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	EFS
	OFS	1.1334	1.1352	1.1309	1.1296	1.1317	1.1313	1.0959	1.0976	1.0936	1.1005	1.0975	1.0996	OFS
	ODI	62/13	61/13	62/14	62/13	61/13	66/13	48/5	47/5	47/5	47/6	47/6	48/6	ODI
	EFI	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	EFI
	OFI	1.1563	1.1541	1.1558	1.1560	1.1547	1.1622	1.0039	1.0018	1.0052	1.0016	1.0020	1.0057	OFI
	ES	.1500	.3000	.4500	.6000	.7500	.9000	.1500	.3000	.4500	.6000	.7500	.9000	ES
	OS	.1519	.3020	.4518	.6007	.7513	.9009	.1453	.2954	.4457	.5964	.7466	.8981	OS
	SDS	.1697	.1639	.1536	.1374	.1134	.0750	.1658	.1601	.1498	.1342	.1110	.0734	SDS
	SS	.0199	.0273	.0188	.0260	.0246	.0078	.0258	.0279	.0306	.0325	.0363	.0261	SS
	KS	.0398	.0640	.0650	.0417	.0606	.0317	.0477	.0379	.0471	.0353	.0336	.0631	KS
	EI	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	EI
	OI	42.3662	34.8610	27.3766	19.9333	12.4139	4.9421	42.7333	35.2259	27.7082	20.1759	12.6641	5.0936	O
	SDI	6.0155	5.8031	5.4323	4.8624	4.0217	2.6494	5.8427	5.6629	5.3309	4.7969	3.9929	2.6651	SDI
39	SI	.0576	.0719	.0717	.0765	.0852	.0909	-.0387	-.0577	-.0646	-.0846	-.0842	-.0769	SI
	KI	-.1227	-.1092	-.1001	-.1115	-.0855	-.0925	.1102	.1084	.1072	.0987	.0626	.1171	KI
	ODS	66/13	67/13	65/13	64/14	60/13	65/13	50/12	55/12	50/12	50/10	52/11	49/10	ODS
	EFS	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	EFS
	OFS	1.1323	1.1355	1.1355	1.1335	1.1294	1.1316	1.0325	1.0345	1.0285	1.0334	1.0318	1.0306	OFS
	ODI	53/6	56/6	55/5	55/5	52/6	57/6	50/15	49/15	53/13	49/12	51/14	53/12	ODI
	EFI	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	EFI
	OFI	1.0350	1.0346	1.0306	1.0324	1.0318	1.0344	1.0777	1.0754	1.0775	1.0730	1.0821	1.0775	OFS

TABLE 3

Some Characteristics of the Sampling Distributions of Slope, Intercept and Critical Statistics of Slope and Intercept for X Values Rectangularly Distributed but with the Y Values Normally Distributed for Every Value of X for Random and Fixed Values of X for Six Values of Standard Error of Estimates and Three Sample Sizes for Two Groups

		Values of X randomly obtained by random sampling from a rectangular distribution						Values of X Fixed but rectangularly distributed						
		Values of Standard Error of Estimate Based on the Correlations Shown												
Number of Cases Per Group		.1500	.3000	.4500	.6000	.7500	.9000	.1500	.3000	.4500	.6000	.7500	.9000	
	ES	.1500	.3000	.4500	.6000	.7500	.9000	.1500	.3000	.4500	.6000	.7500	.9000	ES
	OS	.1405	.2907	.4416	.5928	.7432	.8956	.1383	.2874	.4393	.5896	.7407	.8946	OS
	SDS	.5897	.5689	.5321	.4775	.3950	.2601	.6259	.6025	.5645	.5062	.4185	.2765	SDS
	SS	-.2375	-.2497	-.2337	-.2277	-.2520	-.2353	-.0095	-.0138	-.0118	-.0095	-.0147	-.0197	SS
	KS	4.0254	4.0701	3.8650	3.8441	3.9075	3.7639	.0271	.0475	.0497	.0409	.0314	.0299	KS
	EI	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	42.5000	35.0000	27.5000	20.0000	12.5000	5 0000	EI
	OI	43.1174	35.5908	28.0116	20.4436	12.8652	5.2187	43.1091	35.6402	28.0178	20.4927	12.9293	5.2300	OL
	SDI	18.2942	17.8290	16.9234	15.2492	12.6304	8.3346	22.0984	21.4167	20.0859	18.0170	14.9150	9.9152	SDI
5	SI	.0042	.0517	.1262	.1650	.1700	.1800	-.1236	-.1140	-.1275	-.1307	-.1284	-.1155	SI
	KI	.8346	.9454	1.1326	1.2181	1.2659	1.3124	-.1259	-.1104	-.1235	-.1376	-.1444	-.1452	KI
	ODS	47/10	46/10	47/10	44/9	45/9	45/7	57/8	55/9	57/9	56/10	56/8	55/11	ODS
	EFS	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	EFS
	OFS	1.3531	1.3484	1.3600	1.3557	1.3479	1.3330	1.5638	1.5517	1.5458	1.5528	1.5758	1.5371	OFS
	ODI	60/9	59/10	57/10	60/9	45/9	58/12	54/8	55/10	53/8	56/9	55/8	55/11	ODI
	EFI	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	EFI
	OFI	1.4458	1.4485	1.4411	1.4455	1.4577	1.4332	1.4750	1.4741	1.4610	1.4746	1.4754	1.4683	OFI
	ES	.1500	.3000	.4500	.6000	.7500	.9000	.1500	.3000	.4500	.6000	.7500	.9000	ES
	OS	.1408	.2907	.4420	.5926	.7431	.8959	.1465	.2966	.4468	.5973	.7476	.8985	OS
	SDS	.2965	.2858	.2674	.2397	.1982	.1311	.1466	.1414	.1324	.1186	.0982	.0648	SDS
	SS	.1959	.1943	.1972	.1984	.1907	.1961	.0446	.0424	.0416	.0386	.0456	.0497	SS
	KS	.8752	.8722	.8945	.8754	.8561	.7978	.0813	.0826	.0828	.1022	.0907	.0751	KS
	EI	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	EI
	OI	43.0970	35.6074	28.0461	20.5041	12.9583	5.2844	42.6626	35.1568	27.6463	20.1179	12.6018	5.0643	OI
	SDI	10.6147	10.2615	9.6078	8.6545	7.1966	4.7626	5.1921	5.0663	4.8197	4.4102	3.7493	2.5429	SDI
13	SI	-.0935	-.1012	-.0942	-.1084	-.0955	-.0729	-.0768	-.0800	-.0852	-.0815	-.0682	-.0354	SI
	KI	.4325	.4116	.3991	.4056	.3783	.3202	.0037	-.0188	-.0131	-.0349	-.0346	-.0639	KI
	ODS	50/11	47/12	48/13	49/12	51/13	51/10	46/13	48/13	48/12	44/13	47/13	49/12	ODS
	EFS	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	1.1000	EFS
	OFS	1.0860	1.0843	1.0870	1.0832	1.0825	1.0914	1.1031	1.0991	1.1024	1.1021	1.1062	1.1042	OFS
	ODI	54/12	56/12	55/11	56/13	55/12	55/13	46/4	46/4	44/4	48/4	46/5	47/4	ODI
	EFI	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	1.0952	EFI
	OFI	1.1183	1.1169	1.1183	1.1220	1.1153	1.1086	1.0004	.9973	1.0020	.9974	1.0025	.9966	OFI
	ES	.1500	.3000	.4500	.6000	.7500	.9000	.1500	.3000	.4500	.6000	.7500	.9000	ES
	OS	.1416	.2921	.4425	.5929	.7444	.8964	.1501	.3002	.4501	.6000	.7501	.9001	OS
	SDS	.1589	.1532	.1437	.1285	.1066	.0702	.0838	.0809	.0757	.0677	.0560	.0371	SDS
	SS	.0484	.0472	.0433	.0478	.0557	.0443	.0514	.0484	.0505	.0561	.0456	.0457	SS
	KS	.0867	.0798	.0866	.0947	.0742	.0799	.1213	.1239	.1283	.1557	.1306	.1104	KS
	EI	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	42.5000	35.0000	27.5000	20.0000	12.5000	5.0000	EI
	OI	42.8595	35.3439	27.8319	20.3266	12.7598	5.1749	42.4935	34.9880	27.4946	19.9932	12.4891	4.9907	OI
39	SDI	5.7854	5.5845	5.2334	4.6858	3.8821	2.5670	2.8665	2.7989	2.6594	2.4353	2.0687	1.4141	SDI
	SI	-.1139	-.1118	-.1179	-.1202	-.1226	-.1253	-.0603	-.0624	-.0669	-.0724	-.0575	-.0432	SI
	KI	.1228	.1273	.1475	.1484	.1226	.1279	-.0743	-.0693	-.0762	-.0671	-.0674	-.0789	KI
	ODS	41/6	43/6	39/6	41/7	42/6	47/6	58/16	60/16	58/16	60/16	59/16	57/15	ODS
	EFS	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	1.0278	EFS
	OFS	.9731	.9687	.9709	.9666	.9759	.9687	1.0857	1.0865	1.0860	1.0847	1.0847	1.0808	OFS
	ODI	47/9	47/7	47/8	48/7	50/9	47/9	50/14	51/14	53/15	49/15	51/15	52/13	ODI
	EFI	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274	EFI
	OFI	.9822	.9828	.9867	.9827	.9844	.9754	1.0758	1.0787	1.0775	1.0738	1.0774	1.0798	OFI

47

TABLE 4

VALUES AND MEANS FOR THE DISCREPANCY BETWEEN THE EXPECTED AND OBSERVED VALUES FOR SLOPE CONSIDERING POSSIBLE EFFECTS DUE TO DISTRIBUTION SHAPE, SAMPLING PROCEDURES, STANDARD ERROR OF ESTIMATE AND SAMPLE SIZE

SAMPLING PROCEDURE

RANDOM / FIXED

VALUES OF STANDARD ERROR OF ESTIMATE BASED ON THE CORRELATIONS SHOWN

DISTRIBUTION SHAPE		Random .1500	Random .3000	Random .4500	Random .6000	Random .7500	Random .9000	Random MEAN FOR SAMPLE SIZE	Fixed .1500	Fixed .3000	Fixed .4500	Fixed .6000	Fixed .7500	Fixed .9000	Fixed MEAN FOR SAMPLE SIZE
NORMAL	SAMPLE SIZE 5	-.0378	-.0366	-.0340	-.0315	-.0249	-.0164	-.0302	.0055	.0038	.0050	.0042	.0033	.0027	.0041
	SAMPLE SIZE 13	-.0039	-.0035	-.0038	-.0036	-.0020	-.0017	-.0031	.0144	.0138	.0134	.0148	.0095	.0063	.0120
	SAMPLE SIZE 39	-.0019	-.0020	-.0018	-.0007	-.0013	-.0009	-.0014	.0047	.0046	.0043	.0036	.0034	.0019	.0037
	MEAN FOR STANDARD ERROR OF ESTIMATE	-.0145	-.0140	-.0132	-.0119	-.0094	-.0063		.0082	.0074	.0076	.0075	.0054	.0036	
	MEAN FOR NORMAL DISTRIBUTION							-.0116							.0066
RECTANGULAR	SAMPLE SIZE 5	.0095	.0093	.0084	.0072	.0068	.0044	.0076	.0117	.0126	.0107	.0104	.0093	.0054	.0100
	SAMPLE SIZE 13	.0092	.0093	.0080	.0074	.0069	.0041	.0075	.0035	.0034	.0032	.0027	.0024	.0015	.0028
	SAMPLE SIZE 39	.0084	.0079	.0075	.0071	.0056	.0036	.0067	-.0001	-.0002	-.0001	.0000	-.0001	-.0001	-.0001
	MEAN FOR STANDARD ERROR OF ESTIMATE	.0090	.0088	.0080	.0072	.0064	.0040		.0050	.0053	.0046	.0044	.0039	.0023	
	MEAN FOR RECTANGULAR DISTRIBUTION							.0073							.0042

TABLE 5

VALUES AND MEANS FOR THE DISCREPANCY BETWEEN THE EXPECTED AND OBSERVED VALUES OF INTERCEPT CONSIDERING POSSIBLE EFFECTS DUE TO DISTRIBUTION SHAPE, SAMPLING PROCEDURES, STANDARD ERROR OF ESTIMATE AND SAMPLE SIZE

SAMPLING PROCEDURE

VALUES OF STANDARD ERROR OF ESTIMATE BASED ON THE CORRELATIONS SHOWN

DISTRIBUTION SHAPE		RANDOM .1500	RANDOM .3000	RANDOM .4500	RANDOM .6000	RANDOM .7500	RANDOM .9000	RANDOM MEAN FOR SAMPLE SIZE	FIXED .1500	FIXED .3000	FIXED .4500	FIXED .6000	FIXED .7500	FIXED .9000	FIXED MEAN FOR SAMPLE SIZE
NORMAL	SAMPLE SIZE 5	.4086	.4001	.3411	.3509	.2556	.1709	.3212	-.1794	-.0857	-.1191	-.1024	-.0754	-.0713	-.1055
	13	.1605	.1317	.1321	.1221	.0592	.0504	.1093	-.7188	-.6872	-.6610	-.5819	-.4608	-.3049	-.5691
	39	.1338	.1390	.1[illegible]	.0667	.0861	.0579	.1012	-.2333	-.2259	-.2082	-.1759	-.1641	-.0936	-.1835
	MEAN FOR STANDARD ERROR OF ESTIMATE	.2343	.2236	.1989	.1799	.1336	.0931		-.3772	-.3329	-.3294	-.2867	-.2334	-.1566	
	MEAN FOR NORMAL DISTRIBUTION							.1772							-.2860
RECTANGULAR	SAMPLE SIZE 5	-.6174	-.5908	-.5116	-.4436	-.3652	-.2187	-.4579	-.6091	-.6402	-.5178	-.4927	-.4293	-.2300	-.4865
	13	-.5970	-.6074	-.5461	-.5041	-.4583	-.2844	-.4996	-.1626	-.1568	-.1463	-.1179	-.1018	-.0643	-.1249
	39	-.3595	-.3439	-.3319	-.3266	-.2598	-.1749	-.2994	.0065	.0120	.0054	.0068	.0109	.0093	.0085
	MEAN FOR STANDARD ERROR OF ESTIMATE	-.5246	-.5140	-.4632	-.4248	-.3611	-.2260		-.2551	-.2617	-.2196	-.2013	.1734	.0950	
	MEAN FOR RECTANGULAR DISTRIBUTION							-.4190							-.2010

TABLE 6

VALUES AND MEANS FOR NUMBER OF OBSERVED F VALUES SIGNIFICANT AT THE .05/.01 LEVEL BASED ON DIFFERENCES IN SLOPE CONCERNING POSSIBLE EFFECTS DUE TO DISTRIBUTION SHAPE, SAMPLING PROCEDURE, STANDARD ERROR OF ESTIMATE AND SAMPLE SIZE

SAMPLING PROCEDURE

RANDOM / FIXED

VALUES OF STANDARD ERROR OF ESTIMATE BASED ON THE CORRELATIONS SHOWN

DISTRIBUTION SHAPE	SAMPLE SIZE	RANDOM .1500	RANDOM .3000	RANDOM .4500	RANDOM .6000	RANDOM .7500	RANDOM .9000	RANDOM MEAN FOR SAMPLE SIZE	FIXED .1500	FIXED .3000	FIXED .4500	FIXED .6000	FIXED .7500	FIXED .9000	FIXED MEAN FOR SAMPLE SIZE
NORMAL	5	55/7	52/8	51/8	53/8	51/9	53/8	52.50/8.00	48/11	51/11	51/11	49/11	54/12	51/12	50.67/11.33
	13	53/11	50/10	55/10	57/9	60/10	54/11	54.83/10.17	53/6	52/5	52/6	56/7	52/6	53/6	53.00/6.00
	39	66/13	67/13	65/13	64/14	60/13	65/13	64.50/13.17	50/12	55/12	50/12	50/10	52/11	49/10	51.00/11.17
	MEAN FOR STANDARD ERROR OF ESTIMATE	58.00/10.33	56.33/10.33	57.00/10.33	58.00/10.33	57.00/10.67	57.33/10.67		50.33/9.67	52.67/9.33	51.00/9.67	51.67/9.33	52.67/9.67	51.00/9.33	
	MEAN FOR NORMAL DISTRIBUTION							57.28/10.44							51.56/9.50
RECTANGULAR	5	47/10	46/10	47/10	44/9	45/9	45/7	45.67/9.17	57/18	55/9	57/9	56/10	56/8	55/11	56.00/10.83
	13	50/10	47/12	48/13	49/12	51/13	51/10	49.33/11.67	46/13	48/13	48/12	44/13	47/13	49/12	47.00/12.67
	39	41/6	43/6	39/6	41/7	42/6	47/6	42.17/6.17	58/16	60/16	58/16	60/16	59/16	57/15	58.67/15.83
	MEAN FOR STANDARD ERROR OF ESTIMATE	46.00/8.67	45.33/9.33	44.67/9.67	44.67/9.33	46.00/9.33	47.67/7.67		53.67/15.67	54.33/12.67	54.33/12.33	53.33/13.00	54.00/12.33	53.67/12.67	
	MEAN FOR RECTANGULAP DISTRIBUTION							45.72/9.00							53.89/13.11

TABLE 7

VALUES AND MEANS FOR NUMBER OF OBSERVED F VALUES SIGNIFICANT AT THE .05/.01 LEVEL BASED ON DIFFERENCES IN INTERCEPT CONCERNING POSSIBLE EFFECTS DUE TO DISTRIBUTION SHAPE, SAMPLING PROCEDURE, STANDARD ERROR OF ESTIMATE AND SAMPLE SIZE

SAMPLING PROCEDURE

RANDOM | FIXED

VALUES OF STANDARD ERROR OF ESTIMATE BASED ON THE CORRELATIONS SHOWN

DISTRIBUTION SHAPE	SAMPLE SIZE	Random .1500	Random .3000	Random .4500	Random .6000	Random .7500	Random .9000	Random MEAN FOR SAMPLE SIZE	Fixed .1500	Fixed .3000	Fixed .4500	Fixed .6000	Fixed .7500	Fixed .9000	Fixed MEAN FOR SAMPLE SIZE
NORMAL	5	56/14	52/13	51/13	56/13	56/13	51/13	53.67/13.17	58/10	60/11	59/11	58/10	61/10	59/10	59.17/10.33
	13	62/13	61/13	62/14	62/13	61/13	66/13	62.33/13.17	48/5	47/5	47/5	47/6	47/6	48/6	47.33/5.50
	39	53/6	56/6	55/5	55/5	52/6	57/6	54.67/5.67	50/15	49/15	53/13	49/12	51/14	53/12	50.83/13.50
	MEAN FOR STANDARD ERROR OF ESTIMATE	57.00/11.00	56.33/10.67	56.00/10.67	57.67/10.33	56.33/10.67	58.00/10.67		52.00/10.00	52.00/10.33	53.00/9.67	51.33/9.33	53.00/10.00	53.33/9.33	
	MEAN FOR NORMAL DISTRIBUTION							56.89/10.67							52.44/9.78
RECTANGULAR	5	60/9	59/10	57/10	60/9	45/9	58/12	56.5/9.83	54/8	55/10	53/8	56/9	55/8	55/11	54.67/9.00
	13	54/12	56/12	55/11	56/13	55/12	55/13	55.17/12.17	46/4	46/4	44/4	48/4	46/5	47/4	46.17/4.17
	39	47/9	47/7	47/8	48/7	50/9	47/9	47.67/8.17	50/14	51/14	53/15	49/15	51/15	52/12	51.00/14.17
	MEAN FOR STANDARD ERROR OF ESTIMATE	53.67/10.00	54.00/9.67	53.00/9.67	54.67/9.67	50.00/10.00	53.33/11.33		50.00/8.67	50.67/9.33	50.00/9.00	51.00/9.33	50.67/9.33	51.33/9.00	
	MEAN FOR NORMAL DISTRIBUTION							53.11/10.06							50.61/9.1

and expected mean values of the distributions of intercept and slope and the number of observed $\underline{F}$ values based on differences in intercept and differences in slope significant at the .05 and .01 levels.

Although there usually appears to be some discrepancy between what is observed and what is expected in Tables 2 through 7, these discrepancies are generally of a small magnitude. Some results detectable in these tables are outlined below.

A. Effects on the moments of the distributions of intercept and slope

1. Discrepancies between what is expected and observed for the mean values of the distributions of slope shown in Tables 2, 3 and 4

a. Distribution shape and sampling procedure appear to be related to the discrepancy of the slope in a complex manner. The random normal case tends to over-estimate the population slope with the highest absolute discrepancy while the random rectangular, fixed normal and fixed rectangular cases tend to underestimate the population slope in the above order of increasing absolute discrepancy.

b. Sample size appears to be related to discrepancy for slope with smaller absolute discrepancy associated with larger sample size except for the fixed normal case for sample size thirteen.

c. Standard error of estimate appears to be related to discrepancy for slope with least absolute discrepancy occurring with smaller values of standard error of estimate (higher correlation).

2. Discrepancies between what is expected and observed for the mean values of the distributions of intercept shown in Tables 2, 3 and 5.

a. Distribution shape and sampling procedure appear to be related to the discrepancy of the intercepts in a complex manner. The random normal case tends to underestimate the population intercept with the lowest absolute discrepancy while the fixed rectangular, the fixed normal and random rectangular cases tend to overestimate the population intercept in the above order of increasing absolute discrepancy.

b. Sample size appears to be related to the discrepancy for intercepts with smaller absolute discrepancy associated with larger sample size except for the fixed normal case for sample size thirteen.

c. Standard error of estimate appears to be related to discrepancy with least absolute discrepancy occurring with the smaller values of standard error of estimate (higher correlation).

3. Discrepancies between what is expected and observed for the skewness and kurtosis of the distributions of intercept and slope shown in Tables 2 and 3

a. The skewness and kurtosis of the distributions of intercept and slope only appear to differ substantially from what was expected for the random case for sample size five. For this case the skewness and kurtosis appear to be related to the values of standard error of estimate with larger values of skewness and kurtosis associated with lower values of standard error. The same effect also appears to a smaller degree for the random case when the sample size is thirteen.

B. Effects on the discrepancy between the observed and expected mean values of the distributions of $\underline{F}$ values for intercept and slope shown in Tables 2 and 3

1. The discrepancies between what is expected and observed do not appear to be systematically related to

a. Distribution shape or sampling procedure

b. Sample size

c. Value of standard error of estimate

C. The effects on the critical statistics for decisions concerning differences in slope and intercept

1. Discrepancies between what is expected and observed for the number of $\underline{F}$ values based on differences in slope significant

at the .05 and .01 levels shown in Tables 2, 3 and 6

a. Distribution shape and sampling procedure appear to be related to the number of observed _F_ values significant at the .05 and .01 levels in a complex manner. At the .05 level, the order of average highest occurrence and percent of error relative to the expected value of fifty is random normal (15% error), fixed rectangular (8% error), fixed normal (3% error), and random rectangular (9% error). At the .01 level the order of highest average occurrence of the _F_ values and percent error relative to the expected value of ten is fixed rectangular (31% error), random normal (4% error), fixed normal (5% error), and random rectangular (10% error).

b. Standard error of estimate does not appear to be systematically related to the number of significant _F_ values at the .05 and .01 levels.

c. Sample size does not appear to be systematically related to the number of significant _F_ values at the .05 and .01 levels.

2. Discrepancies between what is expected and observed for the number of _F_ values based on differences in intercept significant at the .05 and .01 levels shown in Tables 2, 3 and 7

a. Distribution shape and sampling procedure appear to be related to the number of observed _F_ values significant at the .05 and .01 levels in the following manner. At the .05 level, the order of average highest occurrence and percent error relative to the expected value of fifty is random normal (14% error), random rectangular (6% error), fixed normal (5% error), and fixed rectangular (1% error). At the .01 level, the order of average highest occurrence and percent error relative to the expected value of ten is random normal (7% error), random rectangular (1% error), fixed normal (2% error), and fixed rectangular (9% error).

b. Standard error of estimate does not appear to be systematically related to the number of significant F values at the .05 and .01 levels.

c. Sample size does not appear to be systematically related to the number of significant F values at the .05 and .01 levels.

In summary, one may say that although few of the observed values are exactly the ones expected, generally these differences are of small magnitude. The difficulties which were predicted concerning non-normality of the distributions of intercept and slope for the random case do appear for small sample sizes. However, the critical features of intercept and slope do not appear to be different enough from what is expected to merit concern for practical purposes. Also the lack of normality does not seem to cause a sufficient increase in the number of type one errors for decisions about differences in intercepts and slopes to merit concern for practical purposes.

Conclusions

In the mathematical treatment of linear models, certain of the independent variables are assumed to be fixed variables (Graybill, 1961). When this assumption is made, it can be shown that the computed estimates of the parameters occurring in the models have the desirable characteristics that they are "good" estimates and are normally distributed.

It is often convenient to overlook this assumption when linear models are utilized in a research situation, since the nature of a variable often does not allow the researcher to select cases with particular values of a variable without discarding large amounts of data. This empirical study was undertaken as an attempt to discover the effects produced for the computed statistics and for the decisions made on the basis of a critical ratio concerning differences in these statistics for a particular family of linear models when certain independent variables are not fixed variables.

The assumption that certain independent variables have fixed values can be violated in one or both of two ways: (1) the researcher can fail to pre-select values of the variables which will be found in the data and utilized to estimate the statistics, and/or (2) measurement error can be introduced through the process of observing the values of the independent variables. Berkson (1950) has shown that if for certain linear models the values of the independent variables are only allowed to assume fixed values, the values of these variables can be observed with error without disturbing the mathematically desirable distributional characteristics of the estimates of the unknown parameters. The present study focused on the effects produced for only one of the two cases of violation of this assumption for which no mathematical solutions are available. The case investigated occurs when certain independent variables are not fixed but are observed without error.

The particular family of linear models investigated are a set presented by Bottenberg and Ward (1963, pp. 76-86), who apply them to problems generally approached by the use of the technique of analysis of covariance. The estimable terms in these models represent the estimates of the parameters of intercepts and slopes of group regression lines.

This study investigated the following consequences of the departures from the assumption for persons doing research: (1) the estimates of the slope and intercept parameters are not "good," (2) the distributions of the values of these estimates are non-normal, and/or (3) the number of erroneous decisions concerning differences in the estimates are greater than expected.

Some difficulty was encountered in interpreting the results of this study because differences between what is observed and what is expected may be due to sampling fluctuation or to systematic fluctuation of a subtle nature caused by varying the various factors in the experiments. The results are probably not rigorous enough to please a mathematical statistician. However, these discrepancies were of a small enough magnitude and the values selected for the various factors were probably general enough that the results are of practical value.

In general, the results of this study indicate that the violation of the assumption that the independent variable is fixed does not produce enough disparity between what is expected and what is observed for any of the previously mentioned consequences to be a problem for persons doing research. Thus it seems reasonable to conclude that for this set of linear models, within the confines of the values of the factors selected for this study and for sample sizes not too small (in the neighborhood of thirteen and greater), the effects of the violation of the assumption that the independent variable is fixed present little or no problem for researchers.

REFERENCES

Alexander, H. W. Elements of mathematical statistics. New York: John Wiley and Sons, Inc., 1961.

Berkson, J. Are there two regressions? Journal of the American Statistical Association, Vol. 45, 1950, 164-180.

Bottenberg, R. A., and Ward, J. H. Applied Multiple Linear Regression. Technical Documentary Report PRL-TDR-63-6, 1963.

Calkins, D. S. An empirical investigation of some effects of the violation of the assumption that the covariable in analysis of covariance is a mathematical variable. Unpublished doctoral dissertation, The University of Texas at Austin, 1971.

Fisher, R. A. Statistical methods for research workers. New York: Hafner Publishing Company, Inc., 1958.

Graybill, F. A. An introduction to linear statistical models. Vol. 1. New York: McGraw-Hill Book Company, Inc., 1961.

Laurens, J. (Ed.) User's manual. Austin, Texas: The University of Texas Computation Center, 1969.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill Book Company, 1962.